PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G10L 3/00 — A1

(11) International Publication Number: WO 98/44483

(43) International Publication Date: 8 October 1998 (08.10.98)

(21) International Application Number: PCT/IL98/00145

(22) International Filing Date: 27 March 1998 (27.03.98)

(30) Priority Data:
08/825,344 28 March 1997 (28.03.97) US

(71) Applicant *(for all designated States except US)*: DIGISPEECH (ISRAEL) LTD. [IL/IL]; Hanitzanim Street 15, 46484 Herzliya (IL).

(72) Inventors; and

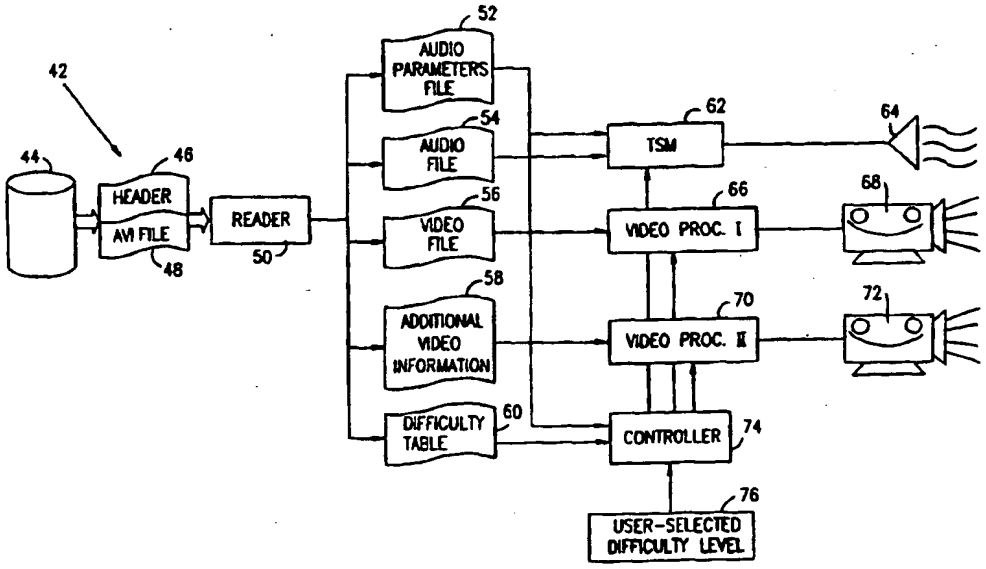
(75) Inventors/Applicants *(for US only)*: SHPIRO, Zeev [IL/IL]; Beit Zuri Street 8, 69122 Tel Aviv (IL). SHAINSKI, Rina [IL/IL]; Nahal Habesor Street 4, 47204 Ramat Hasharon (IL).

(74) Agents: COLB, Sanford, T. et al.; Sanford T. Colb & Co., P.O. Box 2273, 76122 Rehovot (IL).

(81) Designated States: AL, AM, AT, AT (Utility model), AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, CZ (Utility model), DE, DE (Utility model), DK, DK (Utility model), EE, EE (Utility model), ES, FI, FI (Utility model), GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SK (Utility model), SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: TIME SCALE MODIFICATION OF AUDIOVISUAL PLAYBACK AND TEACHING LISTENING COMPREHENSION



(57) Abstract

This invention discloses a digital audiovisual playback system including at least one reader (5) for reading a digital audiovisual memory file (42), a select time base controller (62) receiving an output from the at least one reader, the select time base controller being responsive to a user input for selecting the speed at which audiovisual content read from the digital audiovisual file is played while maintaining audio integrity and synchronization between audio and visual portions of the audiovisual content, and audiovisual output assembly (70) receiving an output from the select time base controller, and providing a user sensible audiovisual output.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

TIME SCALE MODIFICATION OF AUDIOVISUAL PLAYBACK AND TEACHING LISTENING COMPREHENSION

FIELD OF THE INVENTION

The present invention relates generally to audio/video playback and more specifically inter alia to apparatus and methods for learning listening comprehension.

BACKGROUND OF THE INVENTION

Various techniques are known for varying the playback speed of digitally recorded audio-visual materials. Due to difficulties in coordinating the audio portion with the visual portion while maintaining audio playback quality, slow down and speed up functionalities are not commonly provided in audio-visual players.

The present technological limitations on audio-visual playback are also noted in the field of language learning. An example of a relevant recent development in this field is a CD ROM which is distributed free of charge by ALC Press Inc. in Japan in conjunction with their print publication entitled English Network. This CD-ROM teaches listening comprehension by using a video segment taken from a news broadcast and transcribing paragraphs of sentences as they are being spoken.

The following U.S. Patents are believed to be representative of the state of the art: 5,392,163, 5,414,568, 5,418,623, 5,420,801, 5,523,896, 5,543,931, 5,583,652, 5,587,789, 5,596,420, 5,608,582, 5,627,692, 5,664,044, 5,692,092, 5,712,946, and 5,717,828.

SUMMARY OF THE INVENTION

The present invention seeks to provide improved digital audiovisual playback apparatus and methods for providing

for increased or decreased playback speeds while maintaining audio playback quality.

There is thus provided in accordance with a preferred embodiment of the present invention a digital audiovisual playback system including at least one reader for reading a digital audiovisual memory file, a selectable time base controller receiving an output from the at least one reader, the selectable time base controller being responsive to a user input for indicating the speed at which audiovisual content read from the digital audiovisual file is played, while maintaining audio integrity and synchronization between audio and visual portions of the audiovisual content, and an audiovisual output assembly receiving an output from the selectable time base controller and providing a user-sensible audiovisual output.

Further in accordance with a preferred embodiment of the present invention the selectable time base controller is operative to substantially maintain the pitch of the audio portion of the audiovisual memory file notwithstanding changes in the speed at which it is played.

Additionally or alternatively the selectable time base controller is operative to vary time duration of periods of no sound occurring in the audio portion in response to the user input.

Still further in accordance with a preferred embodiment of the present invention the selectable time base controller is operative to vary time duration of periods of sound occurring in the audio portion without substantially altering their pitch.

Additionally in accordance with a preferred embodiment of the present invention the selectable time base controller is operative to synchronize the visual portion with the audio portion.

Still further in accordance with a preferred embodiment of the present invention the selectable time base controller is operative to synchronize the visual portion with the audio portion by either deleting video frames or by repeating or extending presentation or interpolating.

Additionally in accordance with a preferred embodiment of the present invention the selectable time base controller is operative for decreasing the speed of playback of the audiovisual content.

Further in accordance with a preferred embodiment of the present invention the selectable time base controller is operative for increasing the speed of playback of the audiovisual content.

Moreover in accordance with a preferred embodiment of the present invention the selectable time base controller is embodied in a personal computer.

Additionally in accordance with a preferred embodiment of the present invention the selectable time base controller is embodied in a digital video disk player. Alternatively the selectable time base controller is embodied in a dedicated digital video player.

For use in a digital audiovisual playback system, a user-interface controller includes a playback speed selector which enables a user to control playback speed of digital audiovisual content. Preferably the playback speed selector permits a speed variation over a range of at least 200%.

There is also provided in accordance with another preferred embodiment of the present invention a digital audiovisual playback method including the steps of reading a digital audiovisual memory file, selectably controlling playing speed of audiovisual content read from the file by employing a time base controller receiving an output from the at least one

reader, wherein the time base controller, responsive to a user input, selects the speed at which audiovisual content read from the digital audiovisual file is played, while maintaining audio integrity and synchronization between audio and visual portions of the audiovisual content, and receiving an output from the selectable time base controller and providing a user-sensible audiovisual output.

Further in accordance with a preferred embodiment of the present invention the selectable time base controller is operative to substantially maintain the pitch of the audio portion of the audiovisual memory file notwithstanding changes in the speed at which it is played.

Additionally or alternatively the selectable time base controller is operative to vary time duration of periods of non-speech occurring in the audio portion in response to the user input. Preferably the selectable time base controller is operative to vary time duration of periods of speech occurring in the audio portion without substantially altering their pitch. Additionally or alternatively the selectable time base controller is operative to synchronize the visual portion with the audio portion.

Further in accordance with a preferred embodiment of the present invention the selectable time base controller is operative to synchronize the visual portion with the audio portion by either deleting video frames or by repeating or extending existing frames. Preferably the selectable time base controller is operative for decreasing the speed of playback of the audiovisual content. Additionally or alternatively the selectable time base controller is operative for increasing the speed of playback of the audiovisual content.

There is also provided in accordance with another preferred embodiment of the present invention an apparatus for

use in learning listening comprehension including an audio/visual output generator providing synchronized speech and video outputs and a user operable speech output pace controller operative to cause the output generator to provide a speech output at a user selected pace and at a pitch which is generally independent of the selected pace.

Further in accordance with a preferred embodiment of the present invention also including a scorer for sensing user responses and providing a score indication of user achievement level.

Still further in accordance with a preferred embodiment of the present invention the output generator and the controller are operative to provide speech outputs at a pace which is variable over a range of 400 percent.

Additionally in accordance with a preferred embodiment of the present invention the output generator and the controller are operative to provide a speech output whose pace may be varied by both linear and non-linear techniques.

Moreover in accordance with a preferred embodiment of the present invention the scorer is responsive inter alia to the pace at which the speech outputs are provided.

Additionally in accordance with a preferred embodiment of the present invention the video outputs include at least one of images which assist in comprehension of the speech, subtitles and translations.

Preferably the subtitles and translations are synchronized to the pace of the speech outputs.

Further in accordance with a preferred embodiment of the present invention the video outputs include highlighting of portions of the subtitles in synchronization with the speech outputs.

Still further in accordance with a preferred embodiment of the present invention the controller is responsive to a user selected learning level for determining not only the pace of the speech outputs but also whether at least one of subtitles and translations are provided.

Preferably the controller is also responsive to a user selected learning level for determining also whether portions of at least one of subtitles and translations are highlighted in synchronization with said speech outputs.

There is also provided in accordance with yet another preferred embodiment of the present invention a method for teaching listening comprehension including providing an output generator which produces synchronized speech and video outputs, and causing the output generator to provide a speech output at a user selected pace and at a pitch which is generally independent of the selected pace.

Further in accordance with a preferred embodiment of the present invention and also including sensing user responses and providing a score indication of user achievement level.

Still further in accordance with a preferred embodiment of the present invention the speech outputs are provided at a user selectable pace which is variable over a range of 400 percent.

Additionally in accordance with a preferred embodiment of the present invention the speech outputs are provided at a user selectable pace which may be varied by both linear and non-linear techniques.

Moreover in accordance with a preferred embodiment of the present invention the scorer is responsive inter alia to the pace at which the speech outputs are provided.

Still further in accordance with a preferred embodiment of the present invention the video outputs include at

least one of images which assist in comprehension of the speech, subtitles and translations.

Preferably the subtitles and translations are synchronized to the pace of the speech outputs.

Further in accordance with a preferred embodiment of the present invention the video outputs include highlighting of portions of said subtitles in synchronization with the speech outputs.

Still further in accordance with a preferred embodiment of the present invention a user selected learning level determines not only the pace of the speech outputs but also whether at least one of subtitles and translations are provided.

Additionally in accordance with a preferred embodiment of the present invention a user selected learning level determines also whether portions of at least one of subtitles and translations are highlighted in synchronization with the speech outputs.

It is noted that throughout the specification and claims the terms "speech" and "sound" are used interchangeably and refer to spoken words, phrases and sounds as well as non-spoken sounds.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will be more fully understood and appreciated from the following detailed description, taken in conjunction with the drawings in which:

Fig. 1A, 1B 1C, and 1D are illustrations of slowing down an audiovisual playback, Figs. 1A and 1B illustrating the prior art, and Figs. 1C and 1D illustrating a preferred embodiment of the present invention;

Fig. 2A, 2B, 2C, and 2D are illustrations of speeding up an audiovisual playback, Figs. 2A and 2B illustrating the prior art, and Figs. 2C and 2D illustrating a preferred embodiment of the present invention;

Fig. 3 is a block diagram illustration of a digital audiovisual playback system constructed and operative in accordance with a preferred embodiment of the present invention;

Figs. 4A, 4B, and 4C, taken together, are graphical and block diagram illustrations of a preferred mode of operation of the system shown in Fig. 3;

Fig. 5 is a generalized illustration of apparatus for learning listening comprehension constructed and operative in accordance with a preferred embodiment of the present invention;

Fig. 6 is a table illustrating user selectability of various functionalities provided by the apparatus of Fig. 1; and

Fig. 7 is an illustration of a preferred realization of various different audio paces by the apparatus of Fig. 1, while generally maintaining audio pitch uniformity.

DETAILED DESCRIPTION OF A PREFERRED EMBODIMENT

The present invention provides a system and method for selectably, in response to user inputs, slowing down or speeding up audiovisual playback from a digital file. The digital file may be in the form of a digital video tape, a digital video disk, a computer memory, such as a hard disk or a buffer or even a digital memory of a remote server, the contents of which are received concurrently and which may be, but need not necessarily be, stored in a buffer in a client computer.

Reference is now made to Figs. 1A, 1B, 1C and 1D, which illustrate in a simplified manner the operation of the present invention in slowing audiovisual playback in contrast to the prior art. Prior art Fig. 1A illustrates typical original

audiovisual content including a series of continuous video frames 10 and an accompanying audio soundtrack 12, here shown as including speech. It is appreciated that alternatively or additionally, the audio soundtrack 12 may include speech, music, or any other type of sound, and that multiple soundtracks may accompany video frames 10. A time line 14 is shown having several time indices 16 to indicate the passage of time as frames 10 and soundtrack 12 are output.

Fig. 1B shows a prior art technique for slowing down the playback of the frames 10 and soundtrack 12 shown in Fig. 1A. According to the prior art, each frame 10 is played back over a longer time than in the original and the soundtrack 12 is also similarly stretched. This stretching produces a pitch distortion in the audio output which is extremely unpleasant to a user and impairs the integrity of the audio playback, thus decreasing its intelligibility.

In accordance with a preferred embodiment of the present invention, as shown in Figs. 1C and 1D, soundtrack 12 is divided into speech portions 18, representing active audio, and non-speech portions 20, representing the substantially silent intervals between sounds such as between words or phrases. As shown in Fig. 1D, each frame 10 is played back over a longer time than in the original. Soundtrack 12, however, is not stretched to the extent that it is in the prior art. Speech portions 18 may be stretched to a certain extent, such as up to a factor of 2.5, but in a manner which ensures that the pitch is preserved. Furthermore non-speech portions 20 may be increased substantially, as required. Techniques for changing the time basis of speech are described hereinbelow with reference to Figs. 4 - 7.

Furthermore, in accordance with a preferred embodiment of the invention, the audio portion is and continues to be

synchronized with the video portion. This is typically achieved by ensuring that the individual video frames 10 are played substantially over the same time duration as the portion of soundtrack 12 corresponding thereto. If necessary certain video frames may be repeated.

As is shown in Fig. 1D, each speech portion 18 remains synchronized with the video frame to which it originally corresponded, thus maintaining the overall synchronization between the audio and video portions. The factors by which the speech portions 18 and the non-speech portions 20 are stretched are determined and applied in accordance with a difficulty level selected by a user. The video frames are then stretched such that each video frame that has a corresponding speech portion 18 continues to be synchronized with the speech portions 18 to which it originally corresponded.

Reference is now made to Figs. 2A, 2B, 2C, and 2D, which illustrate in a simplified manner the operation of the present invention in speeding up audiovisual playback in contrast to the prior art. Prior art Fig. 2A illustrates typical original audiovisual content including a series of continuous video frames 30 and accompanying audio soundtrack 32, here shown as including speech. It is appreciated that alternatively or additionally, the audio soundtrack 32 may include speech, music, or any other type of sound, and that multiple soundtracks may accompany video frames 30. A time line 34 is shown having several time indices including time index 36 to indicate the passage of time as frames 30 and soundtrack 32 are output.

Fig. 2B shows a prior art technique for speeding up the playback of frames 30 and soundtrack 32 shown in Fig. 2A. According to the prior art, each frame 30 is played back over a shorter time than in the original, and the soundtrack 32 is also similarly speeded up. As seen in Fig. 2B, the frames 30 labeled

'1', '2', and '3', as well as the portion of the soundtrack 34 corresponding to the frames shown, are shown being output partly or completely prior to a time index 36' of a time line 34', with time index 36' corresponding temporally to time index 36 of time line 34 of Fig. 2A. This speeding up produces a pitch distortion in the audio output which is extremely unpleasant to a user and impairs the integrity of the audio playback, thus decreasing its intelligibility.

In accordance with a preferred embodiment of the present invention, as shown in Figs. 2C and 2D, soundtrack 32 is divided into speech portions 38, representing sound such as speech or other active audio, and non-speech portions 40, representing the intervals between words or phrases. As seen in Fig. 2D, the frames 30 labeled '1', '2', and '4', as well as the portion of the soundtrack 34 corresponding to the frames shown, are shown being output partly or completely prior to time index 36' of time line 34', with time index 36' corresponding temporally to time index 36 of time line 34 of Figs. 2A and 2C. Soundtrack 32 is not speeded up to the extent that it is in the prior art. Speech portions 38 may be speeded up to a certain extent, such as up to a factor of 2.5, but in a manner which ensures that the pitch is preserved. Furthermore the non-speech portions 40 may be decreased substantially, as required. Techniques for changing the time basis of speech are described hereinbelow with reference to Fig. 4 - 7.

Furthermore, in accordance with a preferred embodiment of the invention, the audio portion is and continues to be synchronized with the video portion. This is typically achieved by ensuring that the individual video frames 30 are played substantially over the same time duration as the portion of the soundtrack 32 corresponding thereto. If necessary certain video

frames may be discarded, such as the frame 30 labeled '3' is discarded in Fig. 2D.

As is shown in Fig. 2D, each speech portion 38 remains synchronized with the video frame to which it originally corresponded, thus maintaining the overall synchronization between the audio and video portions. The factors by which the speech portions 38 and the non-speech portions 40 are speeded up are determined and applied in accordance with a difficulty level selected by a user. The video frames are then speeded up such that each video frame that has a corresponding speech portion 38 continues to be synchronized with the speech portions 38 to which it originally corresponded.

Reference is now made to Fig. 3 which is a block diagram illustration of a digital audiovisual playback system constructed and operative in accordance with a preferred embodiment of the present invention. A data file 42 including digital audio and video content is typically stored on a storage medium 44 from where it is retrievable. File 42 may comprise a header portion 46, typically containing descriptive information regarding a body portion 48, such as an AVI-format audiovisual file. Header portion 46 typically includes time indices and durations of speech portions corresponding to the audio portion of body portion 48. Header portion 46 may also include data relating to or resulting from TSM pre-processing of body portion 48. Additionally or alternatively, some or all of header portion 46 may be included in a file separate from file 42.

File 42 is typically read at a reader 50 where it is split into audio parameters 52, where audio parameters 52 are typically derived from header 46, an audio portion 54, a video portion 56, and additional video information 58, where additional video information 58 is also typically derived from header 46. A difficulty table 60 is preferably maintained for

controlling audio and video output, as is described in greater detail hereinbelow with reference to Figs. 5 and 6.

A time-scale modifier 62 receives audio parameters 52 and the audio portion 54 and produces a modified audio output 64. A first video processor 66 receives the video portion 56 and produces a video output 68. A second video processor 70 may be used to process the additional video information 58 for use with video processor 66 and/or for additional video output 72. A selectable time base controller 74 preferably controls modifier 62, video processor 66, and video processor 70, referred to collectively as an audiovisual output assembly, to provide a user-sensible audiovisual output. A user interface is preferably provided to receive playback and processing parameters such as a user-selected difficulty level from table 60. The operation of elements of Fig. 3 is described in greater detail hereinbelow with reference to Figs. 4 - 7.

Figs. 4A, 4B, and 4C, taken together, are graphical and block diagram illustrations of a preferred mode of operation of the system shown in Fig. 3. Fig. 4A graphically illustrates audio and video output along a time axis 80. The speed of the video output in Fig. 4A is originally set, for illustration purposes, at 24 frames/second. A video portion 82 is defined as the video frames that correspond to the portion of the audio output that includes actual audio output, in this case speech, while a video portion 84 is defined as the video frames that correspond to the portion of the audio output that does not include speech. The initial duration of video portion 82 and video portion 84 is set, for illustration purposes, at .5 seconds each, with the time elapsed indicated along time axis 80 by a variable t.

A user input is shown in Fig. 4B at 86 as indicating that the video/speech output rate is to be slowed down to .667

of original speed, while the non-speech output rate is to be slowed down to .5 of original speed. As a result, the duration of the speech part increases from .5 seconds to .75 seconds (.5 x 1/.667 seconds = .75 seconds) and the non-speech part from .5 seconds to 1 second (.5 x 1/.5 seconds = 1 second). It has been found through experimentation that adding a non-speech extension, such as the .5 second non-speech extension shown in Fig. 4C, may optimize existing TSM algorithms.

Fig. 4C graphically illustrates audio and video output along time axis 80 as a result of the user input shown in Fig. 4B. As video portion 84 includes 12 frames, the output rate of video portion 82 decreases from 24 frames/second to 16 frames/second in order to accommodate the new speech part duration of .75 seconds. Similarly, the output rate of the remaining 12 frames of video portion 84 decreases from 24 frames/second to 8 frames/second in order to accommodate both the new non-speech part of .1 second as well as the non-speech extension of .5 seconds.

The present invention is particularly suited to applications where digital audiovisual playback is speeded up or slowed down as an aid in research or instruction. For example, the present invention may be implemented as a learning tool to increase listening comprehension as is now described with reference to Figs. 5 - 7.

Reference is now made to Fig. 5, which is a generalized illustration of apparatus for learning listening comprehension constructed and operative in accordance with a preferred embodiment of the present invention. The apparatus of Fig. 5 is preferably embodied in a conventional personal computer 110, such as a Pentium R based personal computer, which is equipped with a keyboard 112, a display 114, a speaker 115, and a mouse 116.

In accordance with a preferred embodiment of the present invention, during learning, the screen of the display 114 appears generally as shown at reference numeral 117 and includes three menu locations 118, 120 and 122, indicated respectively as FILE, DIFFICULTIES, and HELP. A difficulty select scale 124 is also provided for enabling the user to select a level of difficulty, preferably in accordance with a table, such as that illustrated in Fig. 6.

A plurality of operating buttons 126, typically six in number, enable the user to click on one or more of the following typical functionalities: PLAY, STOP, PAUSE/RESUME, SHORT REVERSE, LONG REVERSE, SHORT FORWARD.

A first window 130 illustrates the subject matter of a speech output, which is here indicated at reference numeral 132. A scale 133 may indicate the location of the user in a given lesson and may be used together with a location select functionality thus to enable a user to select a desired location in a lesson.

Additionally, in accordance with a preferred embodiment of the present invention a subtitle 137 may be displayed in a second window, designated by reference numeral 134. This subtitle 137 is preferably a written version of the spoken speech and is synchronized with the spoken speech, as indicated at reference numeral 135. Preferably, a plurality of written words and/or phrases are displayed in window 134 at a given time and the word or phrase currently being spoken is highlighted, as indicated by reference numeral 136.

Further, in accordance with a preferred embodiment of the present invention a translation 142 may be displayed in a third window, designated by reference numeral 138. This translation 142 is also preferably synchronized with the spoken speech. Preferably, a plurality of translated words and/or

phrases are displayed in window 138 at a given time and the word or phrase currently being spoken is highlighted, as indicated by reference numeral 140.

It is a particular feature of the present invention that the timing of the speech output is variable over a relatively wide range, typically up to 400 percent, preferably without appreciably affecting the pitch thereof. In accordance with a preferred embodiment of the invention, as will be described hereinbelow with reference to Fig. 7, both the duration of each word or phrase and the time elapsed between words and/or phrases may be varied. In the speech segment illustrated at reference numeral 135, the speech waveform for each word or phrase is illustrated and its duration is labeled by an index Pn. Intervals between adjacent words and/or phrases are labeled by indices Tn.

Reference is now made to Fig. 6, which is a table illustrating user selectability of various functionalities provided by the apparatus of Fig. 5. It is seen that there are quite a few levels of difficulty, which are distinguished from each other inter alia by one or more of the following:

pace of the speech output which may be expressed in one or both of linear speed of the speech and the amount of pause between words and/or phrases. The amount of pause between words and/or phrases may be varied both by a linear extension and by addition of delay time;

provision of a video output in first window 130;

provision of subtitles in second window 134;

provision of a translation in third window 138; and

synchronized highlighting of the subtitles in second window 134.

Fig. 7 is an illustration of a preferred realization of various different audio paces by the apparatus of Fig. 5,

while generally maintaining audio pitch uniformity. Fig. 7 shows the timing of three different speech output paces, typically as indicated by levels 31 (corresponding to "normal" speech), 11 and 20 in the table of Fig. 6. At the "normal" level, level 31 in the table of Fig. 6, both the duration of each word or phrase and the duration of the interval between each word or phrase are normal for native speakers.

It can be seen that in level 20, both the duration of each word or phrase and the duration of the interval between each word or phrase is extended, albeit by different factors. In level 11 both the duration of each word or phrase and the duration of the interval between each word or phrase are extended, also by different factors, but to an extent greater than in level 20 and an additional pause between each word or phrase is added.

It is to be appreciated that extension of the duration of words and/or phrases and of the duration of the interval between words and/or phrases may be carried out substantially without pitch change by using any suitable algorithm, such as the WSOLA algorithm or the ETSM algorithm. The WSOLA algorithm is described in "An Overlap-Add Technique Based on Waveform Similarity (WSOLA) for High Quality Time-Scale Modification of Speech", ICASSP-93, W. Verhelst and M. Roelands, Vrije Universiteit Brussels, 0-7803-0946-4/93, and the ETSM algorithm is available from Entropic, Cambridge, Massachusetts, USA, Internet address http://www.entropic.com.

It will be appreciated that the present invention is not limited to what has been particularly shown and described hereinabove. Both combinations of various features described herein and subcombinations thereof as well as obvious variations thereof all fall within the scope of the present invention.

C L A I M S

We claim:

1. A digital audiovisual playback system comprising:

at least one reader for reading a digital audiovisual memory file;

a selectable time base controller receiving an output from said at least one reader, said selectable time base controller being responsive to a user input for selecting the speed at which audiovisual content read from the digital audiovisual file is played, while maintaining audio integrity and synchronization between audio and visual portions of said audiovisual content; and

an audiovisual output assembly receiving an output from said selectable time base controller and providing a user-sensible audiovisual output.

2. A digital audiovisual playback system according to claim 1 and wherein said selectable time base controller is operative to substantially maintain the pitch of the audio portion of the audiovisual memory file notwithstanding changes in the speed at which it is played.

3. A digital audiovisual playback system according to claim 1 or claim 2 and wherein said selectable time base controller is operative to vary time duration of periods of no sound occurring in the audio portion in response to said user input.

4. A digital audiovisual playback system according to any of the preceding claims and wherein said selectable time base controller is operative to vary time duration of periods of sound occurring in the audio portion without substantially altering their pitch.

5. A digital audiovisual playback system according to claim 4 and wherein said selectable time base controller is

operative to synchronize the visual portion with the audio portion.

6. A digital audiovisual playback system according to claim 5 and wherein said selectable time base controller is operative to synchronize the visual portion with the audio portion by either deleting video frames or by repeating existing frames.

7. A digital audiovisual playback system according to any of the preceding claims and wherein said selectable time base controller is operative for decreasing the speed of playback of said audiovisual content.

8. A digital audiovisual playback system according to any of the preceding claims and wherein said selectable time base controller is operative for increasing the speed of playback of said audiovisual content.

9. A digital audiovisual playback system according to any of the preceding claims and wherein said selectable time base controller is embodied in a personal computer.

10. A digital audiovisual playback system according to any of claims 1 - 8 and wherein said selectable time base controller is embodied in a digital video disk player.

11. A digital audiovisual playback system according to any of claims 1 - 8 and wherein said selectable time base controller is embodied in a dedicated digital video player.

12. For use in a digital audiovisual playback system according to any of claims 1 - 11, a user-interface controller including a playback speed selector which enables a user to control playback speed of digital audiovisual content.

13. A user-interface controller according to claim 12 and wherein said playback speed selector permits a speed variation over a range of at least 200%.

14. A digital audiovisual playback method comprising the steps of:

reading a digital audiovisual memory file;

selectably controlling playing speed of audiovisual content read from said file by employing a time base controller receiving an output from said at least one reader, wherein said time base controller, responsive to a user input, selects the speed at which audiovisual content read from the digital audiovisual file is played, while maintaining audio integrity and synchronization between audio and visual portions of said audiovisual content; and

receiving an output from said selectable time base controller and providing a user-sensible audiovisual output.

15. A digital audiovisual playback method according to claim 14 and wherein said selectable time base controller is operative to substantially maintain the pitch of the audio portion of the audiovisual memory file notwithstanding changes in the speed at which it is played.

16. A digital audiovisual playback method according to claim 14 or claim 15 and wherein said selectable time base controller is operative to vary time duration of periods of silence occurring in the audio portion in response to said user input.

17. A digital audiovisual playback method according to any of the preceding claims 14 - 16 and wherein said selectable time base controller is operative to vary time duration of periods of sound occurring in the audio portion without substantially altering their pitch.

18. A digital audiovisual playback method according to claim 17 and wherein said selectable time base controller is operative to synchronize the visual portion with the audio portion.

19. A digital audiovisual playback method according to claim 18 and wherein said selectable time base controller is operative to synchronize the visual portion with the audio portion by either deleting video frames or by repeating existing frames.

20. A digital audiovisual playback method according to any of the preceding claims 14 - 19 and wherein said selectable time base controller is operative for decreasing the speed of playback of said audiovisual content.

21. A digital audiovisual playback system according to any of the preceding claims 14 - 20 and wherein said selectable time base controller is operative for increasing the speed of playback of said audiovisual content.

22. Apparatus for use in learning listening comprehension including:

an audio/visual output generator providing synchronized speech and video outputs; and

a user operable speech output pace controller operative to cause the output generator to provide a speech output at a user selected pace and at a pitch which is generally independent of the selected pace.

23. Apparatus according to claim 22 and also comprising:

a scorer for sensing user responses and providing a score indication of user achievement level.

24. Apparatus according to claim 22 and wherein the output generator and said controller are operative to provide speech outputs at a pace which is variable over a range of 400 percent.

25. Apparatus according to claim 22 and wherein said output generator and said controller are operative to provide a speech output whose pace may be varied by both linear and non-linear techniques.

26. Apparatus according to claim 23 and wherein said scorer is responsive inter alia to the pace at which the speech outputs are provided.

27. Apparatus according to claim 22 and wherein said video outputs include at least one of images which assist in comprehension of the speech, subtitles and translations.

28. Apparatus according to claim 27 and wherein said subtitles and translations are synchronized to the pace of the speech outputs.

29. Apparatus according to claim 28 and wherein said video outputs include highlighting of portions of said subtitles in synchronization with said speech outputs.

30. Apparatus according to claim 22 and wherein said controller is responsive to a user selected learning level for determining not only the pace of the speech outputs but also whether at least one of subtitles and translations are provided.

31. Apparatus according to claim 30 and wherein said controller is also responsive to a user selected learning level for determining also whether portions of at least one of subtitles and translations are highlighted in synchronization with said speech outputs.

32. A method for teaching listening comprehension including:

providing an output generator which produces synchronized speech and video outputs;

and causing the output generator to provide a speech output at a user selected pace and at a pitch which is generally independent of the selected pace.

33. A method according to claim 32 and also comprising:

sensing user responses and providing a score indication of user achievement level.

34. A method according to claim 32 and wherein the speech outputs are provided at a user selectable pace which is variable over a range of 400 percent.

35. A method according to claim 32 and wherein said speech outputs are provided at a user selectable pace which may be varied by both linear and non-linear techniques.

36. A method according to claim 33 and wherein said scorer is responsive inter alia to the pace at which the speech outputs are provided.

37. A method according to claim 32 and wherein said video outputs include at least one of images which assist in comprehension of the speech, subtitles and translations.

38. A method according to claim 37 and wherein said subtitles and translations are synchronized to the pace of the speech outputs.

39. A method according to claim 38 and wherein said video outputs include highlighting of portions of said subtitles in synchronization with said speech outputs.

40. A method according to claim 32 and wherein a user selected learning level determines not only the pace of the speech outputs but also whether at least one of subtitles and translations are provided.

41. A method according to claim 40 and wherein a user selected learning level determines also whether portions of at least one of subtitles and translations are highlighted in synchronization with said speech outputs.

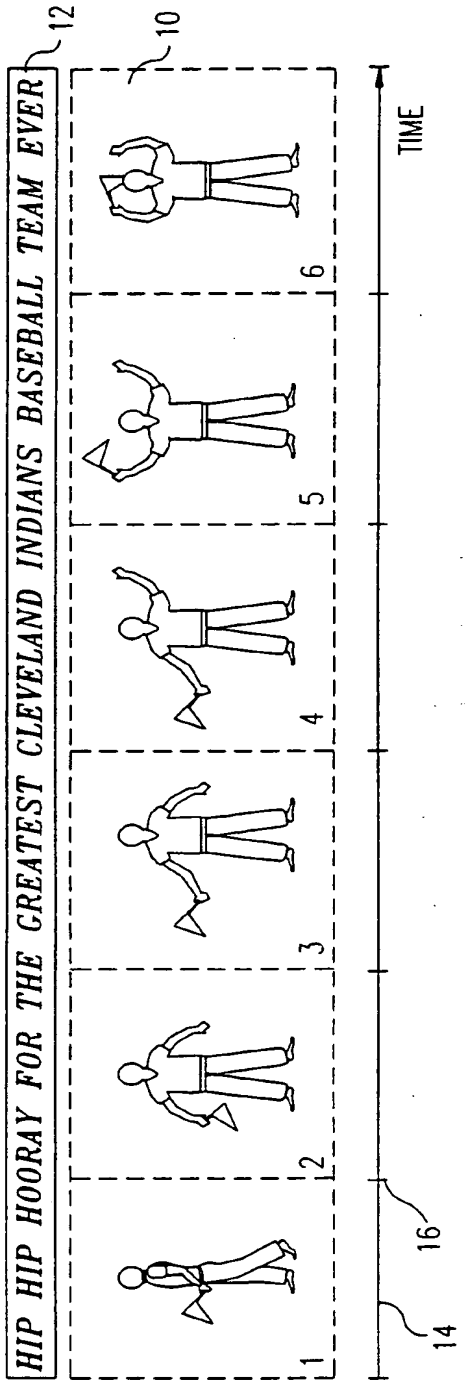


FIG. 1A
PRIOR ART

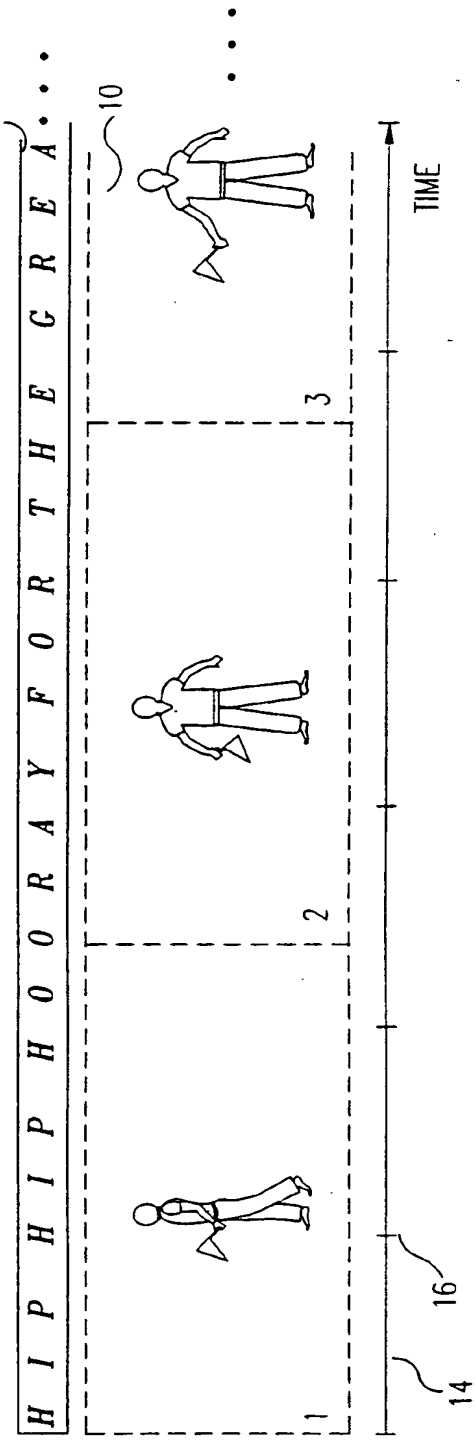


FIG. 1B
PRIOR ART

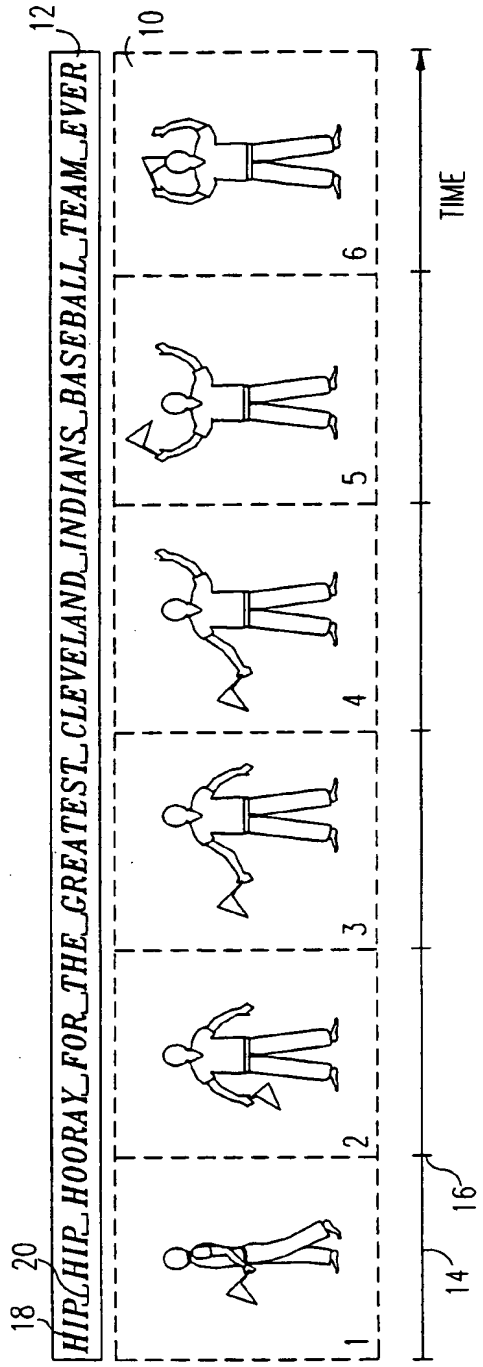

HIP HIP HOORAY FOR THE GREATEST CLEVELAND INDIANS BASEBALL TEAM EVER
12
10
18
20
1
2
3
4
5
6
14
16
TIME

FIG. 1C

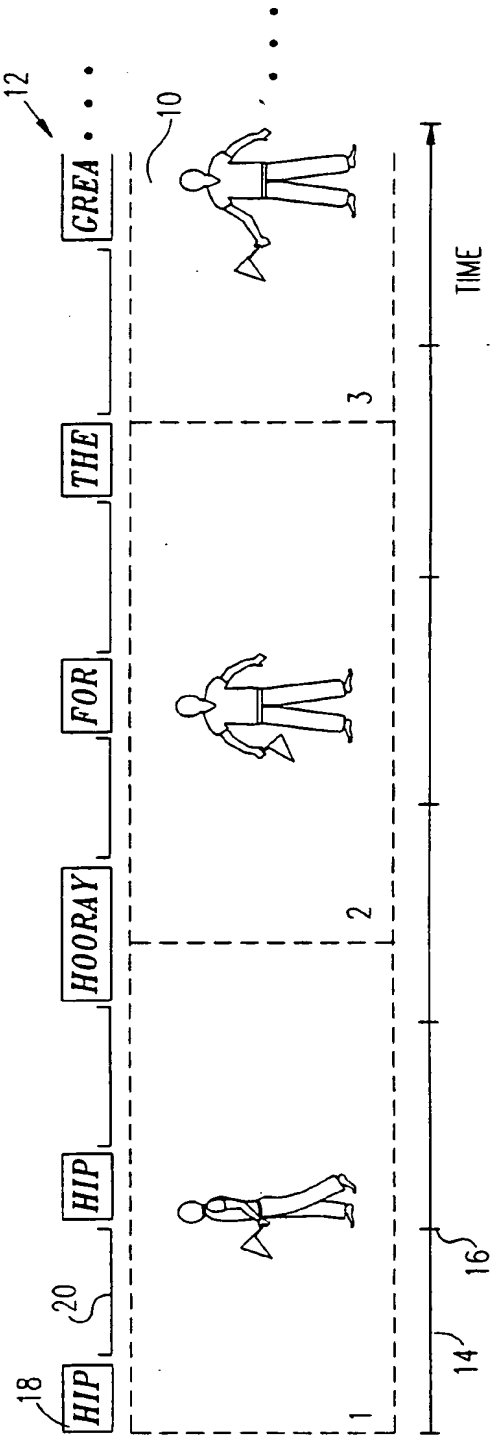

12
HIP
HIP
HOORAY
FOR
THE
CREA . . .
10
18
20
1
2
3
14
16
TIME

FIG. 1D

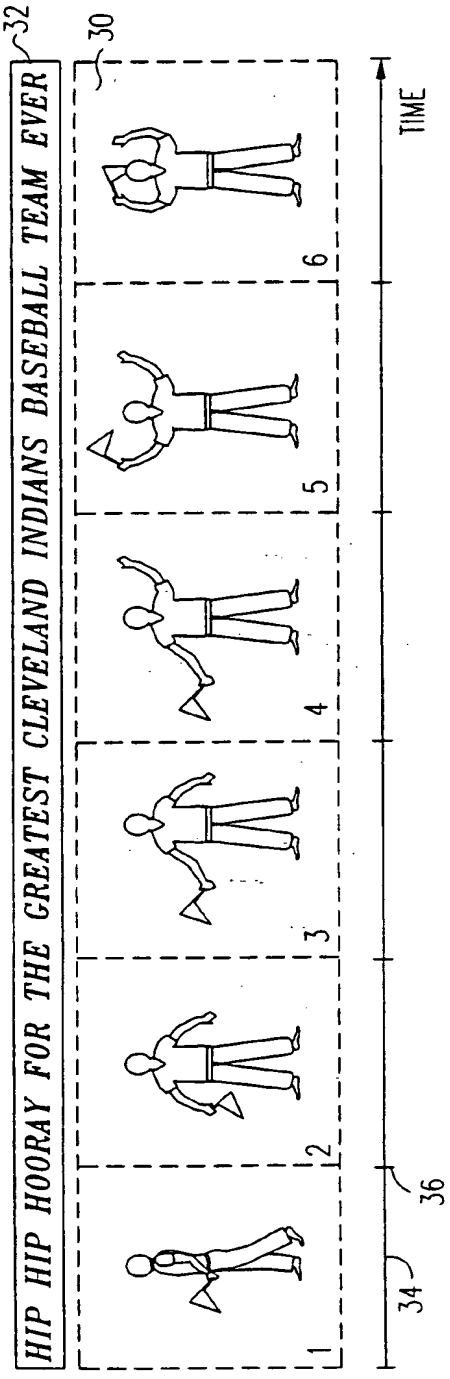


FIG. 2A
PRIOR ART

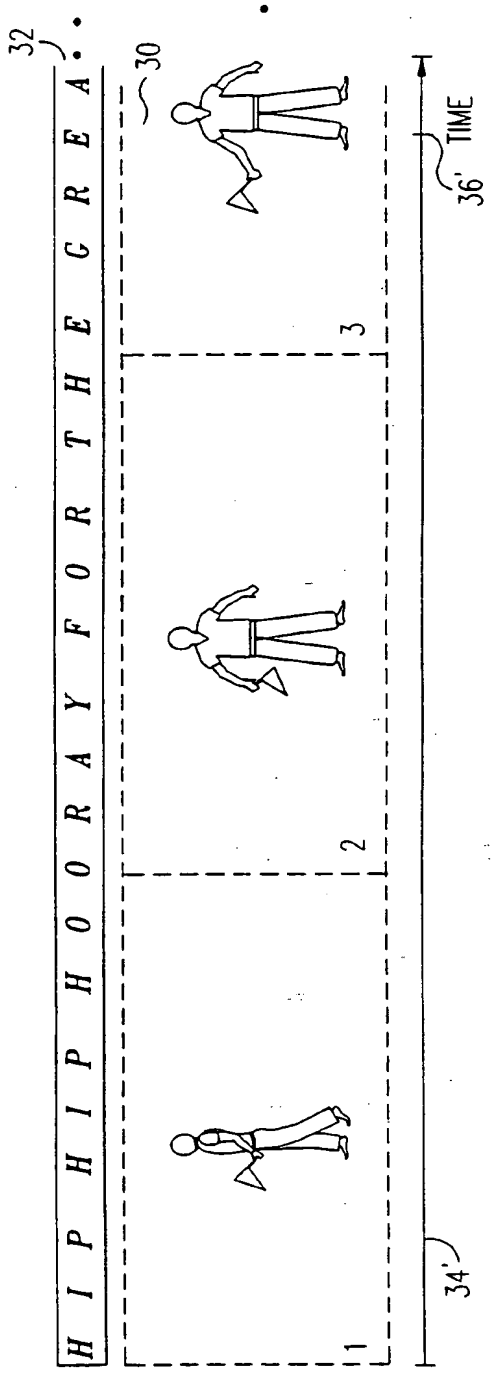


FIG. 2B
PRIOR ART

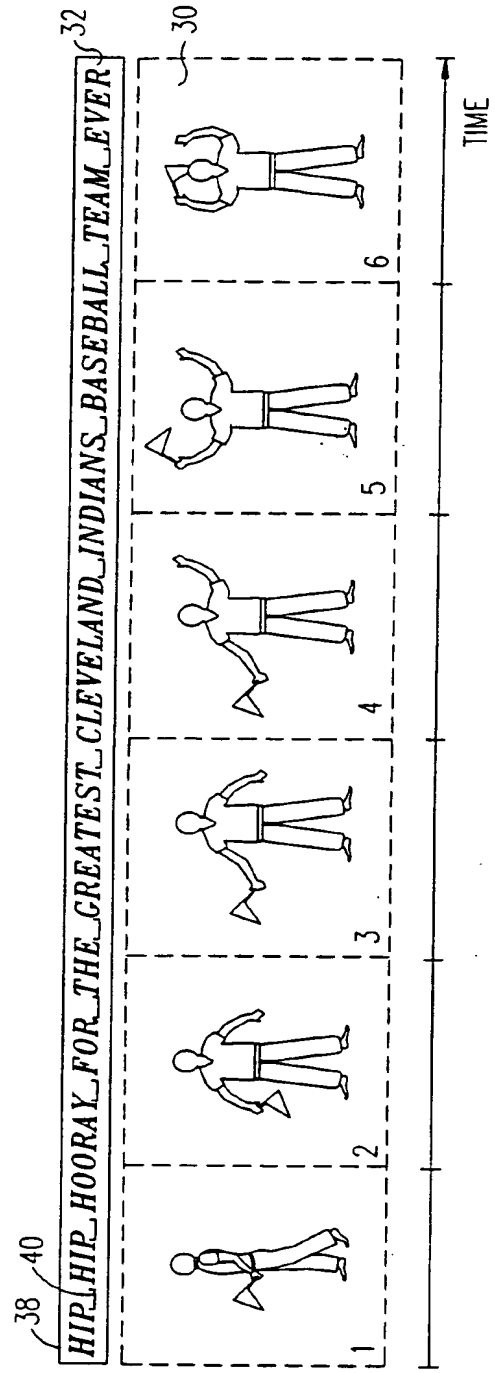

HIP HIP HOORAY FOR THE GREATEST CLEVELAND INDIANS BASEBALL TEAM EVER
38
40
32
30
1
2
3
4
5
6
TIME

FIG. 2C

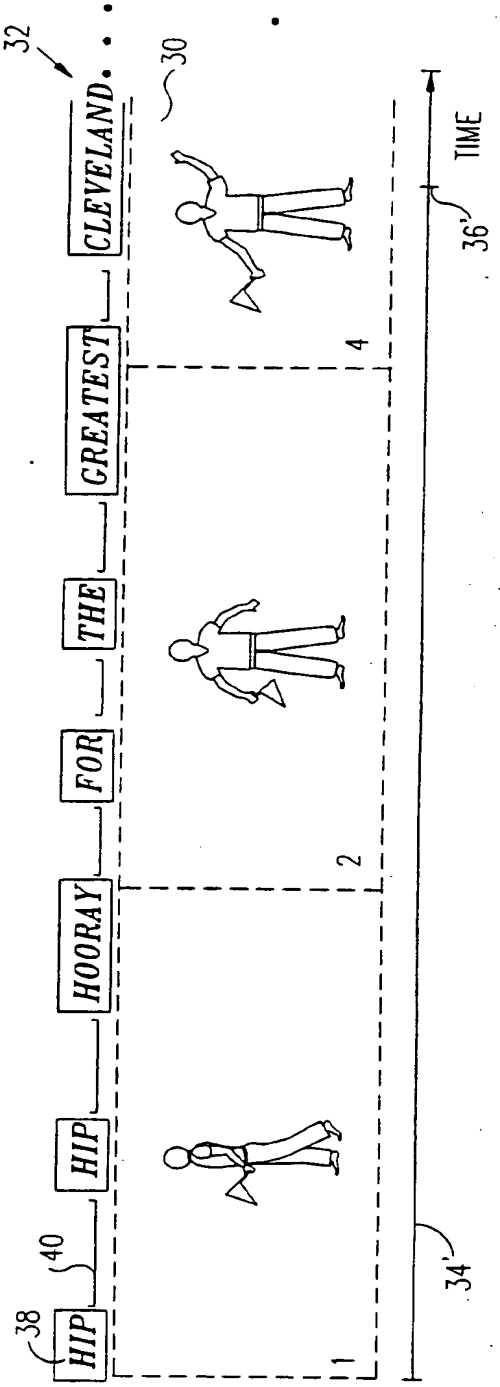

HIP
HIP
HOORAY
FOR
THE
GREATEST
CLEVELAND. . .
38
40
32
30
1
2
4
34'
36'
TIME

FIG. 2D

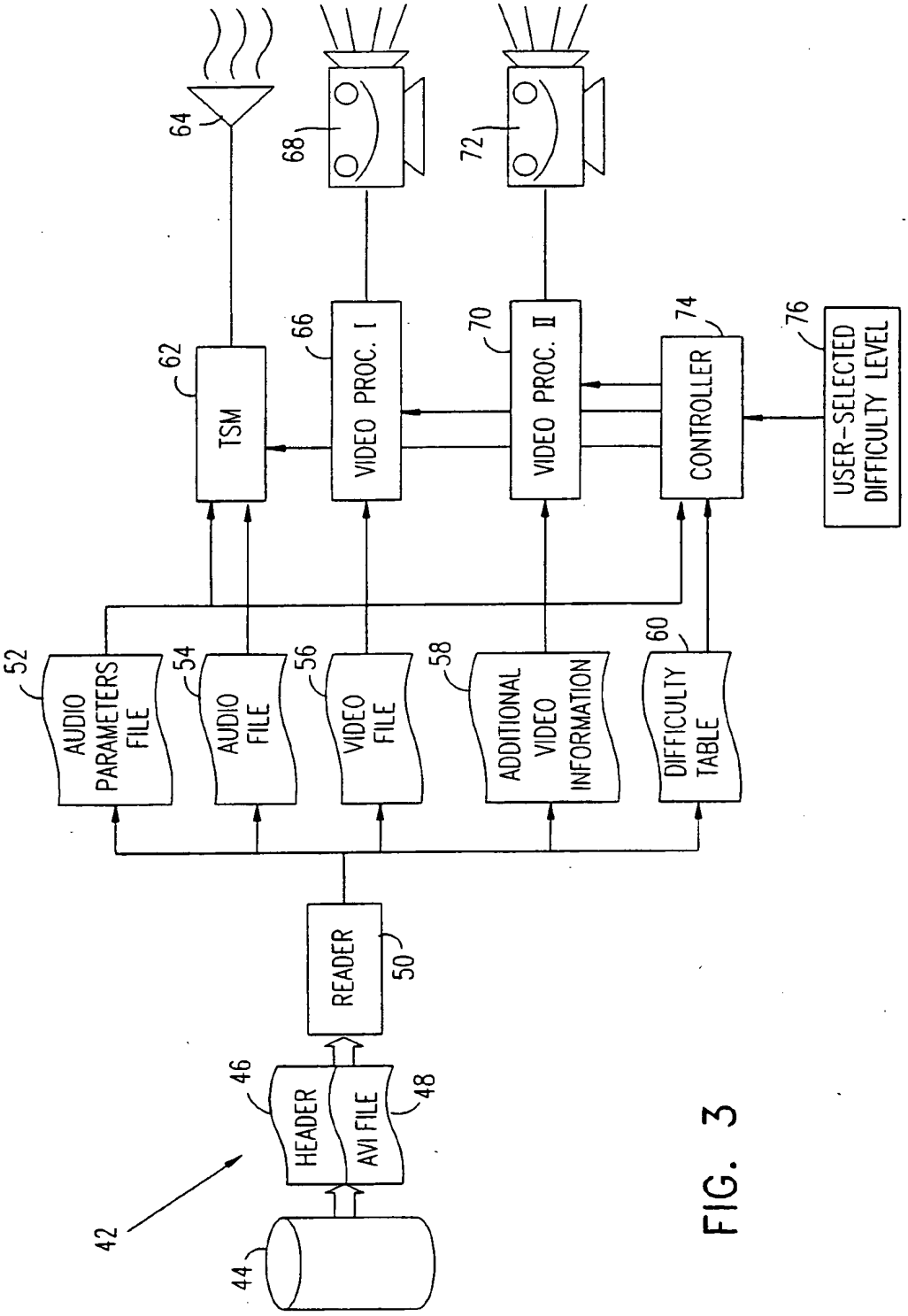

42
44
HEADER
AVI FILE
46
48
READER
50
AUDIO PARAMETERS FILE
52
AUDIO FILE
54
VIDEO FILE
56
ADDITIONAL VIDEO INFORMATION
58
DIFFICULTY TABLE
60
TSM
62
64
VIDEO PROC. I
66
68
VIDEO PROC. II
70
72
CONTROLLER
74
USER-SELECTED DIFFICULTY LEVEL
76

FIG. 3

FIG. 4A

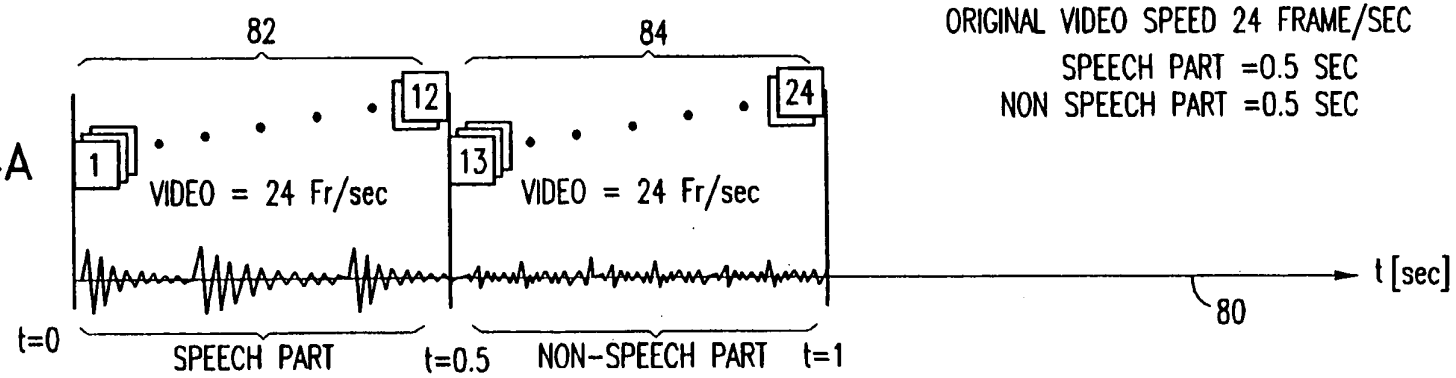


FIG. 4B

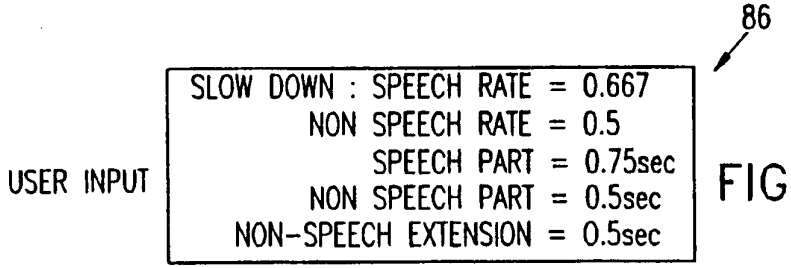


FIG. 4C

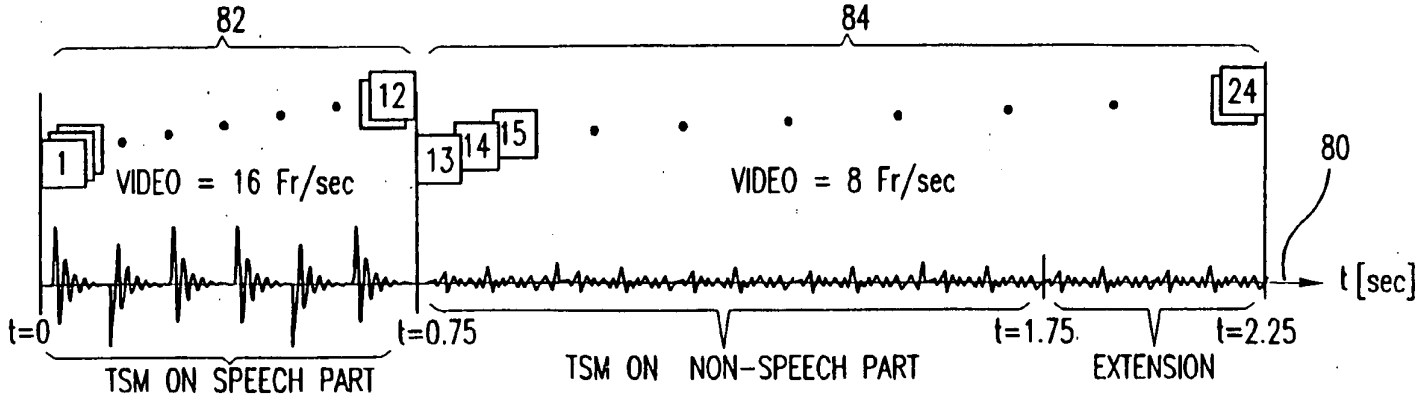
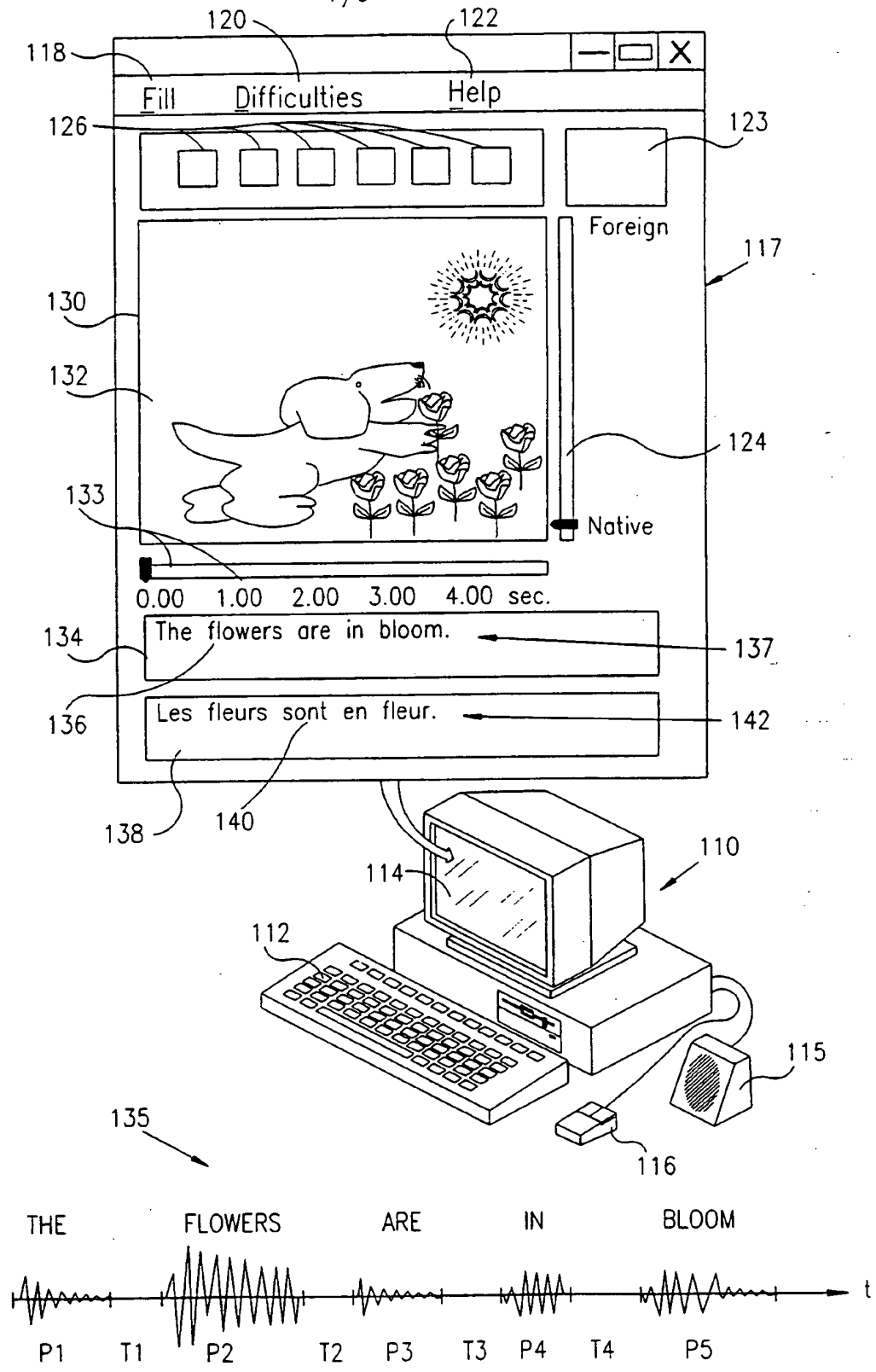


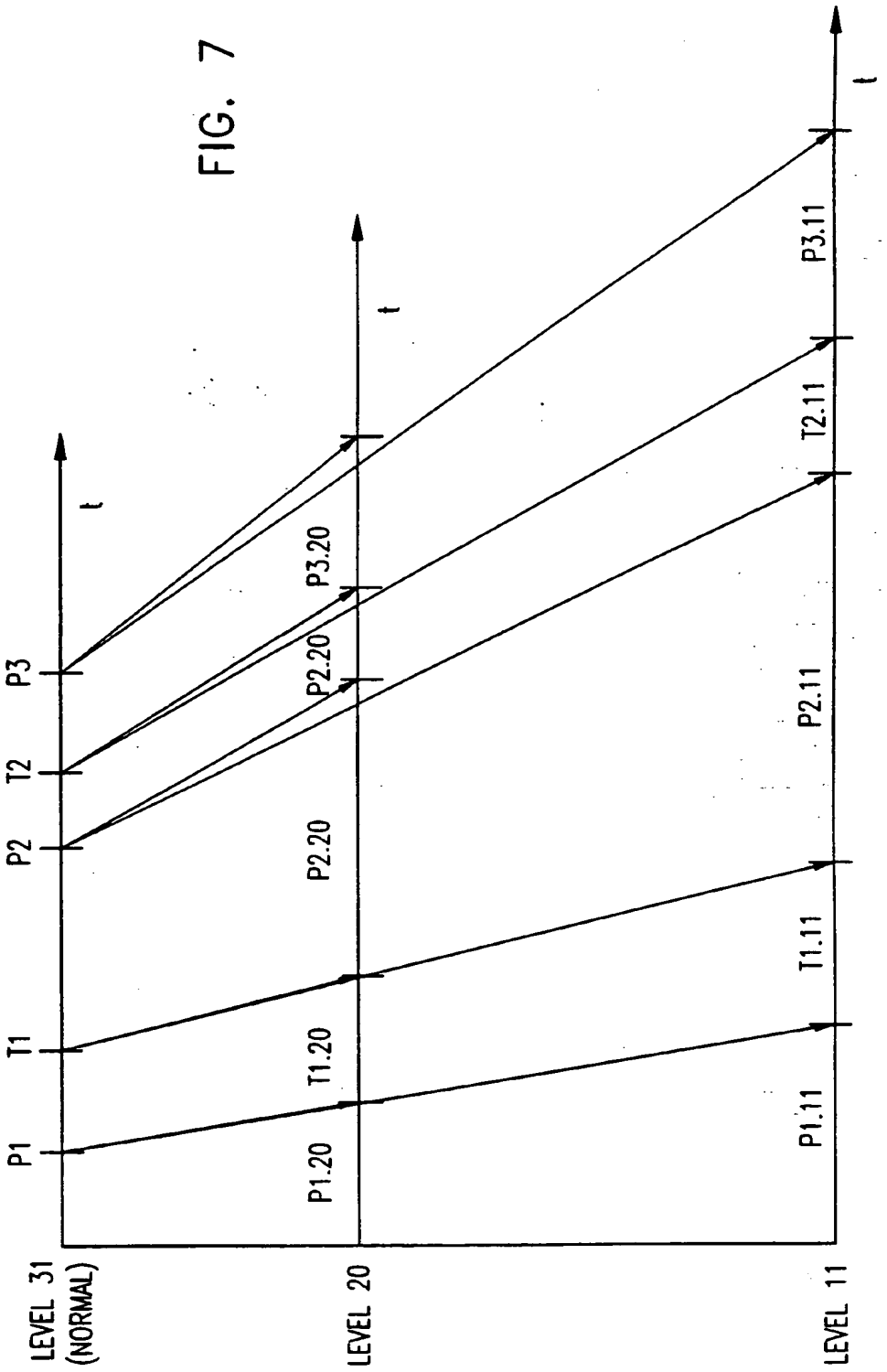
FIG.5


120
122
118
Fill
Difficulties
Help
126
123
Foreign
117
130
132
124
133
Native
0.00 1.00 2.00 3.00 4.00 sec.
134
The flowers are in bloom.
137
136
Les fleurs sont en fleur.
142
138
140
110
114
112
115
135
116
THE
FLOWERS
ARE
IN
BLOOM
t
P1
T1
P2
T2
P3
T3
P4
T4
P5

Level	Phrase Exten.	Pause Exten.	Pause Insert	Video	Eng. Sub.	Pointing	User's Language Sub	Clock
0	2.5	1.5	1sec.	YES	YES	YES	YES	NO
1	2.5	1.5	0.9sec.	YES	YES	YES	YES	NO
2	2.5	1.5	0.8sec.	YES	YES	YES	YES	NO
3	2.5	1.5	0.7sec.	YES	YES	YES	YES	NO
4	2.5	1.5	0.6sec.	YES	YES	YES	YES	NO
5	2.0	1.5	0.6sec.	YES	YES	YES	YES	NO
6	2.0	1.5	0.5sec.	YES	YES	YES	YES	NO
7	2.0	1.5	0.5sec.	YES	YES	YES	YES	NO
8	2.0	1.4	0.4sec.	YES	YES	YES	NO	NO
9	2.0	1.4	0.3sec.	YES	YES	YES	NO	NO
10	2.0	1.4	0.2sec.	YES	YES	YES	NO	NO
11	2.0	1.4	0.1sec.	YES	YES	YES	NO	NO
12	2.0	1.4	NO	YES	YES	YES	NO	NO
13	2.0	1.4	NO	YES	YES	YES	NO	NO
14	2.0	1.3	NO	YES	YES	YES	NO	NO
15	2.0	1.3	NO	YES	YES	YES	NO	NO
16	2.0	1.2	NO	YES	YES	NO	NO	YES
17	2.0	1.25	NO	YES	YES	NO	NO	YES
18	2.0	1.2	NO	YES	YES	NO	NO	YES
19	1.6	1.2	NO	YES	YES	NO	NO	YES
20	1.4	1.15	NO	NO	YES	NO	NO	YES
21	1.3	1.15	NO	NO	YES	NO	NO	YES
22	1.4	1.1	NO	NO	YES	NO	NO	YES
23	1.3	1.1	NO	NO	YES	NO	NO	YES
24	1.4	1.05	NO	NO	YES	NO	NO	YES
25	1.3	1.05	NO	NO	YES	NO	NO	YES
26	1.25	1.0	NO	NO	YES	NO	NO	YES
27	1.25	1.0	NO	NO	YES	NO	NO	YES
28	1.17	1.0	NO	NO	NO	NO	NO	YES
29	1.1	1.0	NO	NO	NO	NO	NO	YES
30	1.05	1.0	NO	NO	NO	NO	NO	YES
31	1.0	1.0	NO	NO	NO	NO	NO	YES

FIG. 6

FIG. 7


t
t
t
P3
T2
P2
T1
P1
P3.20
P2.20
P2.20
T1.20
P1.20
P3.11
T2.11
P2.11
T1.11
P1.11
LEVEL 31
(NORMAL)
LEVEL 20
LEVEL 11

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IL98/00145

A. CLASSIFICATION OF SUBJECT MATTER
IPC(6) :G10L 3/00
US CL :434/185
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 434/185

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A,P	US 5,689,613 A (GASPER ET AL) 18 NOVEMBER 1997, ENTIRE DOCUMENT.	1-41
A	US 4,884,972 A (GASPER) 5 DECEMBER 1989, ENTIRE DOCUMENT.	1-41
A, P	US 5,697,789 A (SAMETH ET AL) 16 DECEMBER 1997, ENTIRE DOCUMENT.	1-41

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 04 JUNE 1998	Date of mailing of the international search report 24 JUN 1998
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer GLENN RICHMAN Telephone No. (703) 308-3170

Form PCT/ISA/210 (second sheet)(July 1992)*